

SAMPLE SOLUTION

1 Concepts and Understanding (F2023.1)

Note that the answers given below are more detailed (for didactical purposes!) than what was actually expected from the students during the examination.

- a. A precise statement of the central-limit theorem is as follows. When N independent random variables x_n are drawn from an arbitrary distribution $p(x)$ with average μ and variance σ^2 , in the limit of large N , their mean X distributes according to a normal (Gaussian) distribution with average μ and variance $N^{-1}\sigma^2$.
- b. With systematic sampling, we have a homogeneous sampling of Ω (in the limit $K \rightarrow \infty$). To obtain the canonical ensemble average A of \mathcal{A} , we must introduce a Boltzmann weighting explicitly, *i.e.*

$$A = Z^{-1} \sum_{k=1}^K \mathcal{A}_k e^{-\beta \mathcal{V}_k} \quad \text{with} \quad Z = \sum_{k=1}^K e^{-\beta \mathcal{V}_k} ,$$

where $\beta = (k_B T)^{-1}$, k_B being the Boltzmann constant and T the absolute temperature. With stochastic sampling, we also have a homogeneous sampling of Ω (in the limit $K \rightarrow \infty$). So, we can use the same equation relying on explicit Boltzmann weighting. With isothermal MD, *i.e.* MD with a thermostat at temperature T , we already sample Ω following an implicit Boltzmann weighting (*i.e.* the vectors \mathbf{r} occur in proportion to their Boltzmann factor). Consequently, the ensemble average is obtained here using plain (arithmetic) averaging, *i.e.*

$$A = K^{-1} \sum_{k=1}^K \mathcal{A}_k .$$

Finally, with umbrella sampling MD, *i.e.* isothermal MD with a biasing potential energy \mathcal{V}_{bias} , weights are also applied implicitly, but they are non-Boltzmannian (Boltzmann factor of $\mathcal{V} + \mathcal{V}_{bias}$ instead of \mathcal{V}). Thus, we must do reweighting when calculating the ensemble average, *i.e.*

$$A = X^{-1} \sum_{k=1}^K \mathcal{A}_k e^{+\beta \mathcal{V}_{bias,k}} \quad \text{with} \quad X = \sum_{k=1}^K e^{+\beta \mathcal{V}_{bias,k}} .$$

- c. A typical potential-energy expression for bond stretching is harmonic, *i.e.*

$$\mathcal{V}_b = \frac{1}{2} k_b (b - b_o)^2 ,$$

where b_o is the reference bond length and k_b the force constant. A typical potential-energy expression for bond-angle bending is harmonic, *i.e.*

$$\mathcal{V}_\theta = \frac{1}{2} k_\theta (\theta - \theta_o)^2 ,$$

where θ_o is the reference bond-angle value and k_θ the force constant. A typical potential-energy expression for torsional-dihedral rotation is a sum of cosine terms, *i.e.*

$$\mathcal{V}_\phi = \sum_m k_{\phi,m} \cos(m\phi - \delta_m) ,$$

where m are multiplicities (positive integers, typically in the range 1-6), while δ_m and k_m are the corresponding phase shifts and force constants, respectively. A typical potential-energy expression for improper-dihedral distortion is harmonic, *i.e.*

$$\mathcal{V}_\xi = \frac{1}{2} k_\xi (\xi - \xi_o)^2 ,$$

where ξ_o is the reference improper-dihedral value and k_ξ the force constant.

2 Fundamental Equations (F2023.2)

Note that the answers given below are more detailed (for didactical purposes!) than what was actually expected from the students during the examination. Watch out not to overlook the requirements “**explain the meaning** of all the involved symbols” and “**state the SI units** of these quantities”.

- a. For a given atom i in the system, the Langevin equation of motion reads

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i^\sim - m_i \gamma_i \dot{\mathbf{r}}_i + \sigma_i \boldsymbol{\eta}_i ,$$

where m_i is the mass (SI unit: kg), $\ddot{\mathbf{r}}_i$ the acceleration (SI unit: m·s⁻²), \mathbf{F}_i^\sim the mean (super-script \sim) force (SI unit: N), γ_i the friction coefficient (SI unit: s⁻¹), $\dot{\mathbf{r}}_i$ the velocity (SI unit: m·s⁻¹), σ_i the stochastic force amplitude (SI unit: N·s^{1/2}), and $\boldsymbol{\eta}_i$ the white-noise vector (SI unit: s^{-1/2}). Commonly, γ_i and σ_i are taken to be identical for all atoms, and simply written as γ and σ . The Langevin equation of motion is typically used in simulations of solutions where the solvent molecules are not included explicitly, *i.e.* when using an implicit-solvent model. It can also be used as a (very effective!) thermostat algorithm when simulating molecules in vacuum, or even in solution with an explicit-solvent representation. The simulation method relying on the integration of the Langevin equation of motion in time is called stochastic dynamics (SD). The left-hand side of the equation corresponds to the mass-acceleration product of the particle, used to propagate its dynamics in time (*e.g.* using the SD leap-frog algorithm). On the right-hand side, the first term corresponds to the mean force on the particle, which encompasses the effect of other solute atoms as well as the mean effect of the solvent. The second term accounts for the friction by the solvent (proportional and in opposite direction to the atomic velocity). Finally, the last term accounts for the random kicks by the solvent molecules, where $\boldsymbol{\eta}_i$ is a random vector sampled from a Wiener (white-noise) process. The generated dynamics is not adiabatic because the friction term removes energy from the system while the stochastic term introduces energy into the system. The balance between the two effects will generate an equilibrium situation. The resulting temperature T depends on the balance between γ (magnitude of the friction) and σ (magnitude of the stochastic kicks) *via* the fluctuation-dissipation equation (assuming here particles with equal masses m)

$$T = \frac{\sigma^2}{2mk_B\gamma} ,$$

where k_B is the Boltzmann constant. In the limit of high friction, the inertial term (left-hand side of the Langevin equation) can be neglected, and one obtains the Brownian equation of motion

$$m_i \gamma_i \dot{\mathbf{r}}_i = \mathbf{F}_i^\sim + \sigma_i \boldsymbol{\eta}_i .$$

The simulation method relying on integrating this equation in time is called Brownian Dynamics (BD).

- b. For a quantum-mechanical system with M energy levels, the canonical partition function is defined by

$$Z = \sum_{m=0}^{M-1} e^{-\beta E_m} ,$$

where E_m is the energy of level m (SI unit: J) with the zero point set as $E_0 = 0$, and $\beta = (k_B T)^{-1}$ (SI unit: J⁻¹), k_B being the Boltzmann constant (SI unit: J·K⁻¹) and T the absolute temperature (SI unit: K). For a classical system of $3N$ particles ($M = 3N$ degrees of freedom), the canonical partition function is defined as

$$Z = \xi \int d\mathbf{x} e^{-\beta \mathcal{H}(\mathbf{x})} ,$$

where \mathbf{x} is the $2M$ -dimensional phase-space vector (coordinate and momenta) of the system (SI unit of the volume element: $\text{J}^M \cdot \text{s}^M$), \mathcal{H} the Hamiltonian (SI unit: J), and $\xi = (h^M N!)^{-1}$ for indistinguishable particles (SI unit: $\text{J}^{-M} \cdot \text{s}^{-M}$), h being the Planck constant (SI unit: J·s). The quantity Z is a central quantity in statistical mechanics because its knowledge (along with that of its derivatives with respect to temperature, volume, and number of particles) provides all the thermodynamic information on the system. For example, it determines the Helmholtz free energy as

$$F = -\beta^{-1} \ln Z + U_o ,$$

where U_o is the zero-point energy, the enthalpy as

$$U = -\frac{\partial \ln Z}{\partial \beta} + U_o ,$$

and the entropy as

$$S = k_B \ln Z + \frac{U - U_o}{T} .$$

3 Derivations (F2023.3)

a. Combining the expression for $W(\mathbf{p})$ with the equation for the Boltzmann entropy, one gets

$$S(\mathbf{p}) = k_B K^{-1} \left[\ln K! - \sum_{m=0}^{M-1} \ln(K p_m)! \right] .$$

Inserting the Stirling approximation $\ln N! \approx N \ln N - N$ gives

$$S(\mathbf{p}) = k_B K^{-1} \left[K \ln K - K - K \sum_{m=0}^{M-1} p_m (\ln p_m + \ln K - 1) \right] .$$

Using the normalization condition (*i.e.* the fact that the sum of the p_m probabilities evaluates to one), this simplifies to

$$S(\mathbf{p}) = -k_B \sum_{m=0}^{M-1} p_m \ln p_m ,$$

which is the Gibbs expression for the entropy.

b. Omitting the zero-point energy (irrelevant here), the equation connecting the Helmholtz free energy F to the canonical partition function Z reads

$$F = -\beta^{-1} \ln Z ,$$

where $\beta = (k_B T)^{-1}$, k_B being the Boltzmann constant and T the absolute temperature. The free-energy difference ΔF_{AB} can thus be written

$$\Delta F_{AB} = -\beta^{-1} \ln \frac{Z_B}{Z_A} .$$

The equation for the classical partition function Z (as determined by the Hamiltonian \mathcal{H}) reads

$$Z = \xi \int d\mathbf{r} d\mathbf{p} e^{-\beta \mathcal{H}} ,$$

with $\xi = (h^{3N} N!)^{-1}$ for N indistinguishable particles, h being the Planck constant. Inserting this into the equation for ΔF_{AB} gives

$$\Delta F_{AB} = -\beta^{-1} \ln \frac{\int d\mathbf{r} d\mathbf{p} e^{-\beta \mathcal{H}_B}}{\int d\mathbf{r} d\mathbf{p} e^{-\beta \mathcal{H}_A}} .$$

This can be re written

$$\Delta F_{AB} = -\beta^{-1} \ln \frac{\int d\mathbf{r} d\mathbf{p} e^{-\beta(\mathcal{H}_B - \mathcal{H}_A)} e^{-\beta \mathcal{H}_A}}{\int d\mathbf{r} d\mathbf{p} e^{-\beta \mathcal{H}_A}} .$$

The expression for the canonical ensemble average $\langle \mathcal{A} \rangle$ of an instantaneous observable \mathcal{A} is

$$\langle \mathcal{A} \rangle = \xi Z^{-1} \int d\mathbf{r} d\mathbf{p} \mathcal{A} e^{-\beta \mathcal{H}} .$$

So we get

$$\Delta F_{AB} = -\beta^{-1} \ln \left\langle e^{-\beta(\mathcal{H}_B - \mathcal{H}_A)} \right\rangle_A ,$$

which is the FEP expression.

4 Explicit Calculations (F2023.4)

- a. First consider the case where the dice have different colors, and a throw is characterized by a sequence. In this situation, the total number of distinguishable sequences for a throw is given by M^K (each of the K dice can have any of M outcomes, and all the resulting combinations are distinguishable). With $M = 6$ and $K = 4$, this evaluates to $6^4 = 1296$. The probabilities of the sequences $\{1, 1, 1, 1\}$, $\{1, 2, 2, 1\}$ and $\{1, 2, 4, 6\}$ are all the same, namely $1/1296$. The most probable sequence(s) is either “any” or “none”, since all are equiprobable. Now considering the case where the dice have the same color, and a throw is characterized by a distribution. In this situation, the total number of distributions for a throw is given by the binomial coefficient

$$N_{dis} = \binom{K + M - 1}{K} = \frac{(K + M - 1)!}{(M - 1)!K!} .$$

With $M = 6$ and $K = 4$, this evaluates to 126. The probability of a distribution $\mathbf{n} = \{n_m \mid m = 1, \dots, M\}$ with $n \in 1, \dots, K$ is given by the multinomial coefficient

$$P(\mathbf{n}) = \binom{K}{n_1, \dots, n_M} = \frac{K!}{\prod_{m=1}^M n_M} .$$

With $K = 4$, one has $K! = 24$. So, the probability of the distribution $\{1, 1, 1, 1, 0, 0\}$ is $24/126$, *i.e.* $4/21$, the probability of the distribution $\{2, 2, 0, 0, 0, 0\}$ is $6/126$, *i.e.* $1/21$, and the probability of the distribution $\{4, 0, 0, 0, 0, 0\}$ is $1/126$. The most probable distributions are those where all the dice show different numbers. There are $M!/((M - K)!K!)$ such combinations, *i.e.* 15, and they all have a probability of $4/21$. You can easily list them.

- b. Using the ideal-gas law, the number n of moles is found to be

$$n = \frac{PV}{RT} = \frac{74.7 \cdot 10^5 \text{ J} \cdot \text{m}^{-3} \cdot 10^{-3} \text{ m}^3}{8.3 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1} \cdot 300 \text{ K}} = \frac{74.7 \cdot 10^2 \text{ J}}{24.9 \cdot 10^2 \text{ J} \cdot \text{mol}^{-1}} = 3 \text{ mol} .$$

There are thus $3 \cdot 6 \cdot 10^{23} = 1.8 \cdot 10^{24}$ molecules in the container. Since the molecules are diatomic, this corresponds to $3.6 \cdot 10^{24}$ atoms.

5 Algorithms and Implementation (F2023.5)

a. A possible code for the function RmsdCalc is given below

```
void RmsdCalc ( int N, int K, double ref[], double trj[][3*N], rmsd[] ) {
    double sum;
    for ( int k = 0; k < K; k++ ) { // loop over trajectory frames
        RotoTransFit(ref,trj[k]); // apply roto-trans fitting (in place)
        sum = 0.0;
        for ( int j = 0; j < 3*N; j++ ) { // loop over atoms and components
            tmp = trj[k][j] - ref[j];
            sum += tmp * tmp;
        }
        rmsd[k] = sqrt(sum/N);
    }
    return;
}
```

The roto-translational superimposition is necessary to eliminate the effect of the rigid-body translation and rotation of the macromolecule, *i.e.* we only want to quantify the structural distortion relative to the reference structure (and not its relative position and orientation). For a simulation performed in solution under periodic boundary conditions (rather than in vacuum), we would also have to restore the connectivity of the macromolecule by periodic gathering (as the periodic copies of the atoms in the reference computational box may not form a “connected” representation of the macromolecule). This must be done prior to the roto-translational fitting.